

Normal-gamma distribution

In probability theory and statistics, the **normal-gamma distribution** (or **Gaussian-gamma distribution**) is a bivariate four-parameter family of continuous probability distributions. It is the conjugate prior of a normal distribution with unknown mean and precision.^[2]

Contents

Definition

Properties

- Probability density function
- Marginal distributions
- Exponential family
- Moments of the natural statistics
- Scaling

Posterior distribution of the parameters

- Interpretation of parameters

Generating normal-gamma random variates

Related distributions

Notes

References

normal-gamma

Parameters	μ location (real) $\lambda > 0$ (real) $\alpha > 0$ (real) $\beta > 0$ (real)
Support	$x \in (-\infty, \infty), \tau \in (0, \infty)$
PDF	$f(x, \tau \mid \mu, \lambda, \alpha, \beta) = \frac{\beta^\alpha \sqrt{\lambda}}{\Gamma(\alpha) \sqrt{2\pi}} \tau^{\alpha - \frac{1}{2}} e^{-\beta \tau} e^{-\frac{\lambda \tau (x - \mu)^2}{2}}$
Mean	^[1] $E(X) = \mu, \quad E(T) = \alpha \beta^{-1}$
Mode	$\left(\mu, \frac{\alpha - \frac{1}{2}}{\beta} \right)$
Variance	^[1] $\text{var}(X) = \left(\frac{\beta}{\lambda(\alpha - 1)} \right), \quad \text{var}(T) = \alpha \beta^{-2}$

Definition

For a pair of random variables, (X, T) , suppose that the conditional distribution of X given T is given by

$$X \mid T \sim N(\mu, 1/(\lambda T)),$$

meaning that the conditional distribution is a normal distribution with mean μ and precision λT — equivalently, with variance $1/(\lambda T)$.

Suppose also that the marginal distribution of T is given by

$$T \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta),$$

where this means that T has a gamma distribution. Here λ , α and β are parameters of the joint distribution.

Then (X, T) has a normal-gamma distribution, and this is denoted by

$$(X, T) \sim \text{NormalGamma}(\mu, \lambda, \alpha, \beta).$$

Properties

Probability density function

The joint probability density function of (X, T) is

$$f(\mathbf{x}, \tau \mid \mu, \lambda, \alpha, \beta) = \frac{\beta^\alpha \sqrt{\lambda}}{\Gamma(\alpha) \sqrt{2\pi}} \tau^{\alpha - \frac{1}{2}} e^{-\beta\tau} \exp\left(-\frac{\lambda\tau(\mathbf{x} - \mu)^2}{2}\right)$$

Marginal distributions

By construction, the marginal distribution of τ is a gamma distribution, and the conditional distribution of \mathbf{x} given τ is a Gaussian distribution. The marginal distribution of \mathbf{x} is a three-parameter non-standardized Student's t-distribution with parameters $(\nu, \mu, \sigma^2) = (2\alpha, \mu, \beta/(\lambda\alpha))$.

Exponential family

The normal-gamma distribution is a four-parameter exponential family with natural parameters $\alpha - 1/2, -\beta - \lambda\mu^2/2, \lambda\mu, -\lambda/2$ and natural statistics $\ln \tau, \tau, \tau\mathbf{x}, \tau\mathbf{x}^2$.

Moments of the natural statistics

The following moments can be easily computed using the moment generating function of the sufficient statistic:

$$\mathbf{E}(\ln T) = \psi(\alpha) - \ln \beta,$$

where $\psi(\alpha)$ is the digamma function,

$$\mathbf{E}(T) = \frac{\alpha}{\beta},$$

$$\mathbf{E}(TX) = \mu \frac{\alpha}{\beta},$$

$$\mathbf{E}(TX^2) = \frac{1}{\lambda} + \mu^2 \frac{\alpha}{\beta}.$$

Scaling

If $(X, T) \sim \text{NormalGamma}(\mu, \lambda, \alpha, \beta)$, then for any $b > 0$, (bX, bT) is distributed as $\text{NormalGamma}(b\mu, \lambda, \alpha, b^2\beta)$.

Posterior distribution of the parameters

Assume that \mathbf{x} is distributed according to a normal distribution with unknown mean μ and precision τ .

$$\mathbf{x} \sim \mathcal{N}(\mu, \tau^{-1})$$

and that the prior distribution on μ and τ , (μ, τ) , has a normal-gamma distribution

$$(\mu, \tau) \sim \text{NormalGamma}(\mu_0, \lambda_0, \alpha_0, \beta_0),$$

for which the density π satisfies

$$\pi(\mu, \tau) \propto \tau^{\alpha_0 - \frac{1}{2}} \exp[-\beta_0\tau] \exp\left[-\frac{\lambda_0\tau(\mu - \mu_0)^2}{2}\right].$$

Suppose

$$\mathbf{x}_1, \dots, \mathbf{x}_n \mid \mu, \tau \sim \text{i.i.d. } \mathcal{N}(\mu, \tau^{-1}),$$

i.e. the components of $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ are conditionally independent given μ, τ and the conditional distribution of each of them given μ, τ is normal with expected value μ and variance $1/\tau$. The posterior distribution of μ and τ given this dataset \mathbf{X} can be analytically determined by Bayes' theorem.^[3] Explicitly,

$$\mathbf{P}(\tau, \mu \mid \mathbf{X}) \propto \mathbf{L}(\mathbf{X} \mid \tau, \mu) \pi(\tau, \mu),$$

where \mathbf{L} is the likelihood of the data given the parameters.

Since the data are i.i.d, the likelihood of the entire dataset is equal to the product of the likelihoods of the individual data samples:

$$\mathbf{L}(\mathbf{X} \mid \tau, \mu) = \prod_{i=1}^n \mathbf{L}(x_i \mid \tau, \mu).$$

This expression can be simplified as follows:

$$\begin{aligned} \mathbf{L}(\mathbf{X} \mid \tau, \mu) &\propto \prod_{i=1}^n \tau^{1/2} \exp\left[-\frac{\tau}{2}(x_i - \mu)^2\right] \\ &\propto \tau^{n/2} \exp\left[-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right] \\ &\propto \tau^{n/2} \exp\left[-\frac{\tau}{2} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2\right] \\ &\propto \tau^{n/2} \exp\left[-\frac{\tau}{2} \sum_{i=1}^n ((x_i - \bar{x})^2 + (\bar{x} - \mu)^2)\right] \\ &\propto \tau^{n/2} \exp\left[-\frac{\tau}{2} (ns + n(\bar{x} - \mu)^2)\right], \end{aligned}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, the mean of the data samples, and $s = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, the sample variance.

The posterior distribution of the parameters is proportional to the prior times the likelihood.

$$\begin{aligned} \mathbf{P}(\tau, \mu \mid \mathbf{X}) &\propto \mathbf{L}(\mathbf{X} \mid \tau, \mu) \pi(\tau, \mu) \\ &\propto \tau^{n/2} \exp\left[-\frac{\tau}{2} (ns + n(\bar{x} - \mu)^2)\right] \tau^{\alpha_0 - \frac{1}{2}} \exp[-\beta_0 \tau] \exp\left[-\frac{\lambda_0 \tau (\mu - \mu_0)^2}{2}\right] \\ &\propto \tau^{\frac{n}{2} + \alpha_0 - \frac{1}{2}} \exp\left[-\tau \left(\frac{1}{2} ns + \beta_0\right)\right] \exp\left[-\frac{\tau}{2} (\lambda_0 (\mu - \mu_0)^2 + n(\bar{x} - \mu)^2)\right] \end{aligned}$$

The final exponential term is simplified by completing the square.

$$\begin{aligned} \lambda_0 (\mu - \mu_0)^2 + n(\bar{x} - \mu)^2 &= \lambda_0 \mu^2 - 2\lambda_0 \mu \mu_0 + \lambda_0 \mu_0^2 + n\mu^2 - 2n\bar{x}\mu + n\bar{x}^2 \\ &= (\lambda_0 + n)\mu^2 - 2(\lambda_0 \mu_0 + n\bar{x})\mu + \lambda_0 \mu_0^2 + n\bar{x}^2 \\ &= (\lambda_0 + n)\left(\mu^2 - 2\frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}\mu\right) + \lambda_0 \mu_0^2 + n\bar{x}^2 \\ &= (\lambda_0 + n)\left(\mu - \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \lambda_0 \mu_0^2 + n\bar{x}^2 - \frac{(\lambda_0 \mu_0 + n\bar{x})^2}{\lambda_0 + n} \\ &= (\lambda_0 + n)\left(\mu - \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \frac{\lambda_0 n(\bar{x} - \mu_0)^2}{\lambda_0 + n} \end{aligned}$$

On inserting this back into the expression above,

$$\begin{aligned} \mathbf{P}(\tau, \mu \mid \mathbf{X}) &\propto \tau^{\frac{n}{2} + \alpha_0 - \frac{1}{2}} \exp\left[-\tau \left(\frac{1}{2} ns + \beta_0\right)\right] \exp\left[-\frac{\tau}{2} \left((\lambda_0 + n) \left(\mu - \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \frac{\lambda_0 n(\bar{x} - \mu_0)^2}{\lambda_0 + n}\right)\right] \\ &\propto \tau^{\frac{n}{2} + \alpha_0 - \frac{1}{2}} \exp\left[-\tau \left(\frac{1}{2} ns + \beta_0 + \frac{\lambda_0 n(\bar{x} - \mu_0)^2}{2(\lambda_0 + n)}\right)\right] \exp\left[-\frac{\tau}{2} (\lambda_0 + n) \left(\mu - \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2\right] \end{aligned}$$

This final expression is in exactly the same form as a Normal-Gamma distribution, i.e.,

$$\mathbf{P}(\tau, \mu | \mathbf{X}) = \text{NormalGamma} \left(\frac{\lambda_0 \mu_0 + n \bar{x}}{\lambda_0 + n}, \lambda_0 + n, \alpha_0 + \frac{n}{2}, \beta_0 + \frac{1}{2} \left(n s + \frac{\lambda_0 n (\bar{x} - \mu_0)^2}{\lambda_0 + n} \right) \right)$$

Interpretation of parameters

The interpretation of parameters in terms of pseudo-observations is as follows:

- The new mean takes a weighted average of the old pseudo-mean and the observed mean, weighted by the number of associated (pseudo-)observations.
- The precision was estimated from 2α pseudo-observations (i.e. possibly a different number of pseudo-observations, to allow the variance of the mean and precision to be controlled separately) with sample mean μ and sample variance $\frac{\beta}{\alpha}$ (i.e. with sum of squared deviations 2β).
- The posterior updates the number of pseudo-observations (λ_0) simply by adding up the corresponding number of new observations (n).
- The new sum of squared deviations is computed by adding the previous respective sums of squared deviations. However, a third "interaction term" is needed because the two sets of squared deviations were computed with respect to different means, and hence the sum of the two underestimates the actual total squared deviation.

As a consequence, if one has a prior mean of μ_0 from n_μ samples and a prior precision of τ_0 from n_τ samples, the prior distribution over μ and τ is

$$\mathbf{P}(\tau, \mu | \mathbf{X}) = \text{NormalGamma} \left(\mu_0, n_\mu, \frac{n_\tau}{2}, \frac{n_\tau}{2\tau_0} \right)$$

and after observing n samples with mean μ and variance s , the posterior probability is

$$\mathbf{P}(\tau, \mu | \mathbf{X}) = \text{NormalGamma} \left(\frac{n_\mu \mu_0 + n \mu}{n_\mu + n}, n_\mu + n, \frac{1}{2} (n_\tau + n), \frac{1}{2} \left(\frac{n_\tau}{\tau_0} + n s + \frac{n_\mu n (\mu - \mu_0)^2}{n_\mu + n} \right) \right)$$

Note that in some programming languages, such as Matlab, the gamma distribution is implemented with the inverse definition of β , so the fourth argument of the Normal-Gamma distribution is $2\tau_0/n_\tau$.

Generating normal-gamma random variates

Generation of random variates is straightforward:

1. Sample τ from a gamma distribution with parameters α and β
2. Sample x from a normal distribution with mean μ and variance $1/(\lambda\tau)$

Related distributions

- The normal-inverse-gamma distribution is essentially the same distribution parameterized by variance rather than precision
- The normal-exponential-gamma distribution

Notes

1. Bernardo & Smith (1993, p. 434)
2. Bernardo & Smith (1993, pages 136, 268, 434)
3. "Archived copy" (<http://www.trinity.edu/cbrown/bayesweb/>). Archived (<https://web.archive.org/web/20140807091855/http://www.trinity.edu/cbrown/bayesweb/>) from the original on 2014-08-07. Retrieved 2014-08-05.

References

- Bernardo, J.M.; Smith, A.F.M. (1993) *Bayesian Theory*, Wiley. ISBN 0-471-49464-X
- Dearden et al. "Bayesian Q-learning" (<http://www.aaai.org/Papers/AAAI/1998/AAAI98-108.pdf>), *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, July 26–30, 1998, Madison, Wisconsin,

USA.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Normal-gamma_distribution&oldid=916436244"

This page was last edited on 18 September 2019, at 22:25 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.